

## **From Field Experiments to Program Implementation: Assessing the Potential Outcomes of an Experimental Intervention Program for Unemployed Persons<sup>1</sup>**

**Amiram D. Vinokur,<sup>2</sup> Richard H. Price, and Robert D. Caplan<sup>3</sup>**

*Michigan Prevention Research Center, Institute for Social Research, The University of Michigan*

*Demonstrated a procedure suggested by Bloom (1984) to provide estimates for the effects of an intervention on its actual participants compared to global effects on study participants in the intervention group, whether or not they showed up. Analyses were based on data collected in a field experiment that tested a preventive intervention for unemployed persons (Caplan, Vinokur, Price, & van Ryn, 1989). Effect size estimates were two to three times larger for the actual participant group than for the entire experimental group on employment outcomes (e.g., earnings) and mental health (anxiety and depression). Further analyses produced results showing that compared to participants, the nonparticipants achieved significantly higher levels of reemployment at posttests and did not differ significantly from participants on all other outcomes. The results suggest that persons who most needed the intervention and benefited from it were drawn into it through self-selection processes.*

In field experiments evaluating preventive interventions, nonparticipation of persons who decline to participate at the outset, or later, by not showing up to the experimental treatment, constitutes a serious threat to the validity of the findings. However, in social intervention programs nonparticipation is a commonplace reality. Since nonparticipation is a pervasive reality in

<sup>1</sup>This article is based on research conducted under NIMH grants #39675 and #2P50MH38330, the latter representing the Michigan Prevention Research Center.

<sup>2</sup>All correspondence should be sent to Amiram D. Vinokur, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48106-1248.

<sup>3</sup>Now at the Department of Psychology, George Washington University.

virtually every type of social intervention program, field experiments that are used to test the intervention for effectiveness should be designed and evaluated with this reality in mind.

Thus, the analyses of randomized field experiments need to be concerned with three research questions, two of which derive from the fact of nonparticipation. First, there is the fundamental question of the overall global effect of the treatment program. Second, there is the question of the effect of the treatment on those who actually participated in the intervention. An intervention program may have an overall low impact due to high no-show and dropout rates, but its treatment may have very strong effects on those who participate. Third, there is the question of what the intervention program might achieve if fully implemented with its target group, that is, if improved recruitment strategies can produce full participation.

It is important to realize that these are each distinct research questions. Furthermore, the answer to these questions may have both theoretical and practical implications for redesigning and implementing tested interventions. For example, if statistically significant global effects were detected for the program, but only weak effects on the participants, the weak effects on the participants would suggest the need to redesign and strengthen the intervention itself. If, however, the effects of the treatment on the actual participants are shown to be strong, then redesigning recruitment and retention strategies would be indicated. And finally, if full participation is insured by newly designed recruitment and retention methods, the effects of the new intervention cannot be extrapolated on the basis of the results of the earlier test that involved partial participation. The reason is that the effects of the intervention on participants with the earlier recruitment methods and incentives may not be the same as those on the type of individuals who chose not to participate then, but now are successfully recruited. Thus, a valid answer to the third question mentioned above can only be obtained through the testing of the new intervention with full participation.

Standard statistical procedures are available for providing a valid answer to the question of the overall effect of the intervention program by comparing the full experimental condition to the control condition. This comparison unequivocally preserves randomization (Cook & Campbell, 1979) and represents a true experimental design. In contrast, answering the latter two questions is limited to the application of specialized methods that depend on the specific conditions of the field experiment.

The purpose of this paper is to explicate the nature of the problems that nonparticipation poses for answering the second and third research questions, for utilizing the findings of field experiments, and for the implementation of the tested intervention in new settings. Furthermore, using

data from a field experiment conducted in our program of prevention research (Caplan, Vinokur, Price, & van Ryn, 1989), we demonstrate how to approach and resolve some of these problems. In particular, we analyze experimental data that include a large percentage of nonparticipants to derive estimates of the intervention's actual effects on those who actually did participate. Finally, we conduct analyses that examine the need for increasing resources to reduce nonparticipation, encouraging a larger proportion of those eligible to take advantage of the program.

The problem posed by nonparticipation of a large proportion of eligible subjects in the experimental test of a social intervention program is twofold. First, nonparticipation seriously threatens the internal validity of the experimental findings. Second, even if the internal validity of the findings is protected, high rates of nonparticipation may drastically restrict the external validity (i.e., the generalizability of the findings). The threat to external validity is a major concern when the levels of participation or the types of persons participating in the experiment are substantially different from what could be achieved when the program is later routinely implemented. In other words, it is not known, and cannot be easily estimated, whether nonparticipants are the type of individuals who would be able to benefit from the intervention as did the original participants. Since the overall effectiveness of an implemented program is dependent on who actually participates, the experimental findings that serve as the basis for the implementation need to be examined with respect to this issue.

The question of how to target the intervention based on who might benefit most can be addressed by comparing the achieved outcomes of the experimental group participants with those of the nonparticipants. If the outcomes of participating individuals are superior to nonparticipants, an argument can be made for the need to increase resources in the implementation phase to reach higher levels of participation. However, if nonparticipants' outcomes are on the same level, or superior to what has been achieved with participants, then it can be argued that the implementation should not aspire to increase rate of participation. Instead, the emphasis should be on using the same methods of recruitment to obtain more participants by having more programs in more locations, but not on intensifying efforts to convince people to participate in order to achieve higher rates of participation in each location.

In the remaining sections we discuss two ways for providing estimates of the experimental effects on the participants in addition to those based on the standard technique of comparing the fully randomized experimental and control conditions irrespective of participation. After demonstrating the application of one of these estimation methods we conduct analyses that address the question of who may benefit more from the intervention.

One way to provide an estimate of intervention effects on participants is based on the construction of a statistical model to predict those who chose to participate and those who declined to participate. Such a predictive model can be constructed and tested using data collected during a pretest from *all* the subjects randomized into the experimental condition. That is, using a multiple regression analysis on the pretest data of the experimental condition, the best predictors of participation can be identified with their respective weights. The model can then be applied to identify the subset of subjects within the control group who would be active participants had they been invited to take part in the intervention (i.e., had they been randomized into the experimental group). More specifically, the multiple regression weights developed from the experimental condition data could be applied to the control condition data to predict the persons most likely to participate in the same proportion as the proportion of participants in the experimental condition. Once this most likely subset of the control group of "would-be participants" is identified, it can be compared with the subgroup of participants in the experimental condition. In the same vein, Heckman (1979) proposed modeling attrition and entering the attrition indicator, assignment to treatment, and their interaction as predictors in a multiple regression equation. The interaction term in this equation represents the net effect of being exposed to the treatment, that is, participation in the intervention program.

The validity of the estimate provided by this method depends on the effectiveness of the modeling; that is, on finding variables that account for a substantial percentage of the variance in participation. In addition, it depends on the extent to which nonparticipants in the experimental group have been treated in the same manner as the control group subjects. If, for example, the control group subjects received an alternative treatment that is hypothesized to produce some effects, this method of estimation cannot be used with any confidence.

Another way to provide a comparison between experimental participants and control group would-be participants is by estimating the latter's contribution of mean score to the total *mean* of the control group using the method suggested by Bloom (1984). Estimating the contribution of the mean of this subgroup to the total mean precludes the need to identify the specific subset of would-be participants in the control group. This method is based on the logic and execution of the randomization procedure, in that the control group is likely to contain the same proportion of persons (with same personal characteristics) who would, if invited to the intervention, be participants as obtained in the experimental group. It is also based on meeting the condition that due to their nonparticipation, the nonparticipants in the experimental group are treated in fundamentally the same way as the control

subjects. If the above conditions hold, according to this method, the mean of the control group's would-be participants can be estimated by subtracting the estimated mean of the would-be nonparticipants from the total mean of the control group. This latter estimate of would-be nonparticipants is based on the *known* mean of the nonparticipants in the experimental group. In other words, the known mean of the actual nonparticipants in the experimental group is substituted for the unknown mean of the would-be nonparticipants in the control group. In this subtraction, the appropriate weights need to be assigned to the various means based on known proportion of such subjects in the experimental group. Specifically, if the proportion of participants in the experimental condition is designed by  $P$ , we can represent the total mean of the control group as:

$$\text{Mean of Control Group} = P * \text{Mean of Would-be Participants} + (1 - P) * \text{Mean of Would-be Nonparticipants}$$

We can then extract the mean of the control group would-be participants from the above formula, and substitute the known mean of the experimental nonparticipants for the unknown mean of the control group would-be nonparticipants. The resulting formula is

$$\text{Mean of Control Would-be Participants} = [P * \text{Mean of Total Control Group} - (1 - P) * \text{Mean of Experimental Nonparticipants}] / P$$

Finally, to obtain the effect of the experimental intervention on the participants we subtract the estimated mean of the control group would-be participants from the mean of the experimental group actual participants. (For more details see Bloom, 1984, pp. 228-229.)

The validity of this second estimation method depends only on the two conditions, which are also essential for the validity of the former method, that is, (a) successfully executed randomization, and (b) the extent to which the nonparticipants have been treated in the same manner as the control group subjects. This method, however, does not require and therefore is not dependent on predictive modeling of participation. Since the statistical modeling of participation in our study did not yield an adequate predictive model, we focus our examination on the application of the second method for estimating the effects of participation as offered by Bloom (1984).

The implications of our discussion and analyses are demonstrated using the data from a study conducted by Caplan et al. (1989). In this study, theories of adherence to difficult courses of action and findings from previous survey research on coping with a major life event—job loss—were

used to generate a preventive intervention that was tested in a randomized field experiment on a broad cross-section sample of the unemployed. The aim was to promote reemployment in high-quality jobs and to prevent poor mental health and loss of motivation to seek reemployment among those who continued to be unemployed.

## METHOD

Since a detailed presentation of the methods is included in Caplan et al. (1989) only brief summaries are provided below.

### *Sample*

*Sites of Recruitment.* Recruitment took place at four offices of state employment compensation offices in southeastern Michigan. Trained interviewers recruited 1,087 persons into the study.

*Characteristics of the Sample.* The sample was intended to represent a broad range of unemployed people. It was similar in some ways to the U.S. unemployed population over 16 years of age (U.S. Bureau of Labor Statistics, 1986) and to representative community survey samples of unemployed persons (e.g., Kessler, Turner, & House, 1988). Males constituted 46% of the sample compared to 60% in the community survey and 56% in the U.S. population. Blacks constituted 15% in our study compared to 20% in the community survey and 22% in the U.S. population. The average age was 35.9 years ( $SD = 10.6$ ) and the average education was about 12.9 years ( $SD = 1.9$ ). Similarly, the average age in the community survey was 35.0 years ( $SD = 10.5$ ) and the average education was about 12.0 years ( $SD = 2.4$ ), as it is in the U.S. population.

*Method of Recruitment.* Respondents were approached while waiting in line at state employment offices and were briefly told about two programs being offered by the University of Michigan on how to seek jobs. One program was described as a 2-week series of morning sessions (the experimental condition); the other was described as a self-guided booklet program (the control condition). Persons were asked if they were interested in participating. Their responses were used to assign them to the conditions listed in Table I.

*Control, Experimental, and Refusers Groups.* Among persons who said they were interested in participating, the interviewer asked whether they preferred the seminar or the self-administered booklet program. To ensure equal motivation to enter one or the other condition, only persons who

Table I. Study Design

Respondent type	Pretest		Posttests <sup>a</sup>			
	Jan.-June 1986		1.5 months		4 months	
	T <sub>1</sub>		T <sub>2</sub>		T <sub>3</sub>	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Control	322	87	281	88	214	67
Experimental	606	81	412	89	414	89
Participants	308	99	282	90	285	92
Dropouts <sup>b</sup>	298	68	130	87	129	86
Refusers	159	84	132	83	132	83
Total <sup>c</sup>	1087	83	825	88	760	81

<sup>a</sup>Of the 1,087 Ss, only 938 were sampled to be included in the posttests. A random half of the T<sub>1</sub> dropouts (149 Ss) were not included in the posttests.

<sup>b</sup>Fifty percent of the sample of dropouts from T<sub>1</sub> were selected for follow-up at T<sub>2</sub> and thereafter. At T<sub>2</sub>, for example, 130/149 dropout respondents (as opposed to 130/298) for a rate of 87%. The percentages are computed as (number received/number of questionnaires mailed at the wave) × 100.

<sup>c</sup>Based on the number eligible participants, dropouts (see Footnote b), and refusers.

expressed no preference were randomly assigned to the experimental and control conditions. Those who expressed a preference were sent job search materials and eliminated from the sample. The majority of these persons preferred the booklet program. In addition to the above two groups, a random sample of 190 unemployed persons who expressed no interest and declined to participate in one program or another, henceforth referred to as refusers, were nevertheless asked to fill out the self-administered questionnaires that were used for data collection to assess the various outcomes.

*Dropouts and Participants.* Among the 752 persons assigned to the experimental condition, 440 (nearly 59%) failed to show up for the intervention and are considered in the analyses as nonparticipants. This percentage varied only by about 5% over the course of recruiting 15 experimental groups in a 4-month period during which successively recruited groups were entered into the experimental and control conditions. "Participants," by contrast, were defined as having completed at least 1 of 8 sessions. The mean number of sessions attended by those who showed up was 6.2 ( $SD = 2.1$ ).<sup>4</sup>

<sup>4</sup>To provide a clear operational definition of nonparticipants, only those who did not show up for the intervention were considered as nonparticipants. The 73 persons who completed between 1 to 5 sessions were included in the participant group. Preliminary analyses showed that the basic findings of the study were not altered by omitting from the analyses the data of these 73 persons.

### *Timing of Data Collection*

Approximately 2 weeks before the intervention began, a self-administered pretest ( $T_1$ ) questionnaire was mailed to all the respondents. The mailed materials included a \$5 bill as payment for completing the questionnaire and a prepaid return envelope. Posttests were, likewise, administered by mail along with similar payments at 4 weeks ( $T_2$ ) and 4 months ( $T_3$ ) after the intervention. Finally, data to monitor the intervention's process were collected at the end of training sessions 1, 7, and 8.

### *Response Rate*

Table I presents the number of respondents and the response rates for the study. Of those experimental, control, and refusers group respondents who received a pretest questionnaire, 83% mailed it back. The response rates for those receiving the  $T_2$  and  $T_3$  posttest questionnaires were 88 and 81% of the preceding  $T_1$  pretest, respectively. The reported analyses were based on the subset of persons who had complete data at all three waves. The response rate varied little from subgroup to subgroup, and as indicated in Table I, the rate for the control group was somewhat lower than those for the experimental group only at  $T_3$ . This  $T_3$  variation, as described in the results, did not contribute to significant pretest treatment differences.

### *Treatment Conditions*

The experimental condition consisted of eight 3-hour training sessions distributed over 2 weeks, 4 mornings per week. All persons in the experimental condition were mailed an advance \$5 incentive to cover transportation costs. Experimental participants were also told that they would receive a \$20 payment for completing at least 6 of the 8 sessions and a certificate of participation.

The design for the eight sessions included the application of problem-solving and decision-making processes, inoculation against setbacks, receiving social support and positive regard from the trainers, and learning and practicing job-seeking skills. The intervention sessions were delivered to groups of 16 to 20 participants by male-female pairs of trainers. The sessions covered a wide range of substantive, skill-related topics. The topics included examples and exercises in identifying and conveying one's job-related skills, using social networks to obtain job leads, contacting potential employers, preparing job application and resumes, and going through a job interview.



The control condition consisted of a booklet briefly describing job-seeking tips equivalent to 2.5 single-spaced pages of text. This booklet was mailed to persons after they were randomized into the control condition. The booklet contained useful information, but it was extremely brief, in comparison to self-help paperback books available on job-seeking.

### *Measures*

In most cases, the major construct assessed in the study consisted of multiitem indices. Most of the resultant measures have coefficient alphas in the .70s and .80s. The major dependent variables included reemployment, job-related variables, and measures of mental health such as anxiety and depression. Other relevant measures included economic hardship and valence of work. In addition, assessments of the immediate perception of the process within the intervention provided an indication of the intervention's integrity and strength (Yeaton & Sechrest, 1981). These assessments were based on a composite measure, "Participant Psychological Engagement."

The job-related measures that were included for the reemployed respondents were indicators of the quality of the job such as monthly earning, quality of work, having a job in one's main occupation, and having a permanent (vs. temporary) job. The measures for the unemployed included motivation to seek reemployment and job-seeking confidence.

In this study, reemployment status was determined by a combination of two measures. To be classified as "reemployed" the person had to report working at least 20 hours per week *and* had to characterize the number of hours employed as "working enough." Persons working less than 20 hours per week *and* characterizing that amount "not working enough" were categorized as "not reemployed." Persons who did not clearly fall into either of these categories (14%) were omitted from analyses that include the reemployment measure. This operational definition provides an unambiguous characterization in that the person is coded as employed only when meeting both subjective and objective criteria.

## RESULTS

### *Effectiveness of Randomization*

To examine the effectiveness of randomization, the experimental and control group were compared on demographic variables, job-seeking mo-

Table II. Means, Effect Sizes, and *t* Tests for Differences Between Full Experimental and Control Groups and Between Experimental Group Actual Participants and Control Group Would-be Participants in Reemployment, Quality of Reemployment, Job-Seeking Confidence, Motivation, Depression, and Anxiety

Outcome variables	Time	Full experimental and control groups				Experimental actual participants and control would-be participants			
		Experimental	Control	Effective size	<i>t</i> ( <i>df</i> )	Actual participants	Would-be part.	Effective size	<i>t</i> ( <i>df</i> )
<b>All subjects</b>									
% Reemployed <sup>a</sup>	T <sub>2</sub>	33	26	0.15	1.70(527)	22	-4	0.60	4.86(319) <sup>e</sup>
	T <sub>3</sub>	60	51	0.17	1.90(532)	53	29	0.48	4.12(338) <sup>e</sup>
	T <sub>2</sub>	1.84	1.91	-0.09	-1.12(630)	1.82	1.97	-0.20	-1.85(383)
Depression	T <sub>3</sub>	1.84	1.92	-0.11	-1.30(623)	1.82	2.01	-0.24	-2.19(381) <sup>c</sup>
	T <sub>2</sub>	1.87	1.88	-0.03	-0.36(630)	1.82	1.84	-0.03	-0.28(383)
Anxiety	T <sub>3</sub>	1.89	1.86	0.04	0.45(622)	1.84	1.74	0.12	1.16(380)
<b>Reemployed subjects</b>									
Monthly Earnings (\$)	T <sub>2</sub>	1461	1084	0.48	2.52(118) <sup>c</sup>	1597	658	1.26	5.58(76) <sup>e</sup>
	T <sub>3</sub>	1465	1372	0.10	0.75(263)	1466	1244	0.23	1.49(175)
Quality of life at work <sup>b</sup>	T <sub>2</sub>	5.02	4.81	0.20	1.15(138)	5.10	4.58	0.51	2.43(88) <sup>c</sup>
	T <sub>3</sub>	4.97	4.75	0.20	1.60(284)	4.94	4.41	0.51	3.37(190) <sup>d</sup>
% in main occupation	T <sub>2</sub>	82	64	0.44	2.54(145) <sup>c</sup>	84	40	1.03	4.98(92) <sup>e</sup>
	T <sub>3</sub>	87	76	0.28	2.28(290) <sup>c</sup>	85	60	0.67	4.42(192) <sup>e</sup>

% in permanent jobs	T <sub>2</sub>	65	53	0.25	1.47(144)	66	36	0.60	2.91(91) <sup>d</sup>
	T <sub>3</sub>	71	64	0.15	1.17(285)	67	49	0.37	2.45(187) <sup>c</sup>
	T <sub>2</sub>	1.73	1.57	0.24	1.41(146)	1.63	1.26	0.63	3.04(92) <sup>d</sup>
Depression	T <sub>3</sub>	1.67	1.78	-0.17	-1.39(289)	1.61	1.85	-0.38	2.52(191) <sup>c</sup>
	T <sub>2</sub>	1.75	1.61	0.23	1.36(146)	1.65	1.30	0.62	3.03(92) <sup>d</sup>
	T <sub>3</sub>	1.71	1.72	0.03	-0.21(289)	1.65	1.66	-0.02	0.16(191)
Unemployed subjects									
Job-seeking confidence	T <sub>2</sub>	3.91	3.77	0.17	1.67(379)	4.00	3.82	0.23	1.42(225)
	T <sub>3</sub>	3.93	3.73	0.25	1.89(231)	3.96	3.50	0.58	2.89(139) <sup>d</sup>
Job-seeking motivation	T <sub>2</sub>	4.92	4.84	0.08	0.73(374)	4.94	4.80	0.13	0.82(223)
	T <sub>3</sub>	4.64	4.38	0.21	1.55(231)	4.59	3.90	0.53	2.73(140) <sup>d</sup>
	T <sub>2</sub>	1.89	2.01	-0.16	-1.52(377)	1.88	2.16	-0.38	-2.33(223) <sup>c</sup>
Depression	T <sub>3</sub>	1.93	2.02	-0.11	-0.86(233)	1.99	2.30	-0.36	-1.84(140)
	T <sub>2</sub>	1.93	1.97	-0.06	-0.54(377)	1.87	1.90	-0.03	-0.22(223)
	T <sub>3</sub>	1.98	1.94	0.04	0.27(232)	1.96	1.87	0.10	0.52(139)

<sup>a</sup>Unemployed and reemployed were coded, respectively, 0 and 1. Tests of significance are based on *t* test of these codes, which produce results identical to test of differences between proportions.

<sup>b</sup>Quality of life at work (index based on 9 major life domains).

<sup>c</sup>*p* < .05.

<sup>d</sup>*p* < .01.

<sup>e</sup>*p* < .001.

tivation, mental health, and other dependent variables assessed at pretest. There were no significant differences on any of these variables.

### *Manipulation Checks, Integrity, and Strength of the Intervention*

A good indication of the integrity and strength of the intervention was provided by the measures of degree of participant engagement. The high mean scores on these measures, ranging from 3.6 to 4.6 on the 5-point scales, suggested that the intervention process produced trust among the participants and that the participants actively practiced skills and dealt with potential setbacks. Furthermore, the participants found both the trainers and the group attractive and supportive. It thus appears that the intervention was perceived as psychologically and socially positive by the participants.

### *Comparing the Intervention Effects in the Original Experimental Design Versus Participation*

Comparison of the effects of the intervention according to the original randomized assignment with the effects based on participation is provided in Table II. The left side of the table contains the means for the full intact experimental and control group, effect sizes (J. Cohen, 1977), and  $t$  tests for the difference between the means. In contrast, the right side of the table contains the data for the experimental subgroup who participated in at least one session, and *estimates* of means and percentages for the control group would-be participants. It also contains effect sizes and  $t$  tests for the mean differences. The results are presented separately for those who became reemployed and for those who remained unemployed at Time 2 and 3 posttests. As discussed above, the estimates for the control group would-be participants were produced using the procedure developed by Bloom (1984).<sup>5</sup>

<sup>5</sup>To compute the  $t$  test statistics for the effects of participation, Bloom suggested using an unbiased estimate of the standard error based on the standard deviations of the control group, the experimental shows, and the experimental noshows with the appropriate weights (Bloom, 1984; p. 229, Equation 6). Because this estimate relies on three parameters and their weights, its own sampling distribution has a large standard deviation; and therefore, it results in an exceedingly conservative and inefficient estimate of the standard error. Consequently, the  $t$  tests based on this estimate produce invariably smaller  $t$  statistics than those based on the entire sample, even when the degrees of freedom are adjusted to be equal. The unbiased estimate is unreasonably conservative: While the difference between the means of our participants and their counterparts is about twice as large as the difference between the mean of the experimental and control groups, the  $t$  statistics for this comparison is much smaller than that for the full groups. To overcome this problem, our calculations of the  $t$  tests are based on the standard formula for the  $t$  test statistics for difference in means

*Effects of the Intervention on Reemployment.* These intervention effects are examined by comparing the effect sizes across the left and right side of Table II. This comparison demonstrates clearly that the effects of the intervention were far more dramatic for the participants than might be inferred from the results based on the full experimental design. For example, regarding the percentage of reemployed, the effect size for the participants at  $T_2$  and  $T_3$  is .60 and .48, which is over three times larger than the .15 and .17 for the full experimental design groups. Among the reemployed, at the  $T_2$  1-month posttest, there was a significantly higher level of earnings in the full experimental than the control group,  $t(118) = 2.52, p < .05$ . By the 4-month posttest, this difference was no longer significant. Furthermore, the percentage of persons who had found reemployment in what they characterized as their main occupation was significantly higher for the experimental group at both  $T_2$  (82% compared to 64%),  $t(145) = 2.54, p < .05$ , and  $T_3$  (87% compared to 76%),  $t(290) = 2.28, p < .05$ . The same results appear in the comparisons for the participants (right side of the table) with consistently larger effect size and  $t$  statistics. In addition, the comparisons for the participants display statistically significant benefits at both posttests with respect to two additional outcomes. Compared to their control counterparts, the reemployed participants enjoyed better quality of life at work,  $t(88) = 2.43, p < .05$  for  $T_2$  and  $t(190) = 3.37, p < .01$  for  $T_3$ , and were more likely to obtain jobs which were characterized as permanent rather than temporary,  $t(91) = 2.91, p < .01$  for  $T_2$  and  $t(187) = 2.45, p < .05$  for  $T_3$ .

*Effects of the Intervention on the Unemployed.* Once again, on the right side of Table II, the effects on job-seeking confidence and motivation of the unemployed group were far more pronounced in the comparisons for the participants with statistically significant effects for these variables at  $T_3$ ,  $t(139) = 2.89, t(140) = 2.73$ , respectively, both  $p > .01$ . In comparison to the would-be participants, the intervention had the effect of maintaining confidence and a sense of efficacy even in the face of setbacks.

---

between the shows in the experimental group and the would-be shows in the control group. The numerator includes Bloom's Equation 4. In the denominator, the unknown standard deviation of the would-be shows in the control group is replaced by the known standard deviation of the entire control group. Similarly, the unknown number of would-be shows in the control group is estimated as the number of cases in the control group multiplied by the proportion of shows in the experimental group. Although our method deviates from the statistical elegance of unbiased estimation, which in this case produces extremely conservative estimates, it provides a more reasonable one. In this study as well as another one conducted by our colleagues (Haney, 1991), each using about a dozen dependent variables, we found that the known variances of the three subgroups, the experimental shows, no-shows, and the controls were similar. In each case, even a conservative estimate of the denominator based on the largest of the three variances was still much smaller than the unbiased estimate as suggested by Bloom.

*Effects on Mental Health.* The comparisons based on the full experimental design did not yield any statistically significant results with respect to mental health indicators such as depression, anxiety, anger, and self-esteem. In contrast, a number of significant differences on these variables, in particular on depression, were revealed in the analyses for the participants. Depression in particular was consistently lower for the unemployed participants at  $T_2$ ,  $t(223) = -2.33$ ,  $p < .05$ , and also  $T_3$ ,  $t(140) = -1.84$ ,  $p < .05$ , one-tail. Surprisingly, the reemployed participants displayed significantly higher level of depression and anxiety at  $T_2$  than their counterpart controls. However, the direction of this difference was reversed at  $T_3$ , with the reemployed participants showing significantly lower levels of depression,  $t(191) = -2.52$ ,  $p < .05$ .

The overall pattern of results is striking in showing greater and more pervasive effects on the actual participants than on the entire experimental group. This is not surprising since over half of the experimental group received no training at all. With respect to the employment variables presented in Table II, 3 of the 14 comparisons based on the experimental design were statistically significant whereas no more than 1 would be expected by chance. In contrast, when the comparisons for the participants are considered, 11 of the 14 comparisons were statistically significant and with larger effect sizes than those yielded by the full experimental design comparisons. Moreover, using the full experimental design comparisons, we did not find significant differences on any of a number of mental health variables. In contrast, using the comparisons for the participants and their counterpart controls, we found that the participants, reemployed and unemployed, reported significantly lower levels of depression at Time 3. Finally, the pattern of results consistently showed that the comparisons based on the full experimental design produced exceedingly conservative estimates for the effects of the intervention on the participants. The effect size estimates were lower by a factor of two to three times than those produced by the comparisons that focused on the participants and their counterparts in the control group.

*Comparisons of Participants with Nonparticipants on the Achieved Outcomes: Who Needs the Intervention Most?*

To examine the question of who needs the intervention most, our analyses shifted to comparisons between those who participated in the intervention and those who declined participation. The latter consisted of two subgroups: those persons from the experimental group who did not show up for the intervention and those who expressed no interest in the

**Table III.** Means, Standard Deviations, and *t* Tests for Differences, Between Experimental Participants and Nonparticipants in Demographics, Quality of Reemployment, Job-Seeking Confidence, and Motivation<sup>a</sup>

Outcome variables	Time	Experimental participants		Experimental nonparticipants		<i>t</i> ( <i>df</i> )
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
All subjects						
Age (in years)	T <sub>1</sub>	40.33	10.67	34.62	10.39	5.72(449) <sup>c</sup>
Sex (M = 1, F = 2)	T <sub>1</sub>	1.53	0.50	1.56	0.50	-0.56(447)
Education	T <sub>1</sub>	13.18	1.88	12.81	1.84	2.11(451) <sup>c</sup>
Income	T <sub>1</sub>	6.13	2.80	5.27	3.12	2.94(411) <sup>d</sup>
Economic hardship	T <sub>1</sub>	2.70	1.14	2.92	1.21	-1.95(443)
Work valence	T <sub>1</sub>	3.99	0.75	3.95	0.71	0.57(448)
Job-seeking motivation	T <sub>1</sub>	4.89	0.96	4.81	1.04	0.79(436)
Job-seeking confidence	T <sub>1</sub>	3.42	0.87	3.83	0.84	-5.06(445) <sup>c</sup>
Quality of past work	T <sub>1</sub>	4.37	0.87	4.36	1.00	0.01(443)
% reemployed <sup>b</sup>	T <sub>2</sub>	19	0.39	37	0.48	-4.04(379) <sup>c</sup>
	T <sub>3</sub>	56	0.50	64	0.48	-1.62(386)
Reemployed subjects						
Monthly earnings (\$)	T <sub>2</sub>	1676	724	1307	733	2.15(90) <sup>c</sup>
	T <sub>3</sub>	1501	778	1411	817	0.81(209)
Unemployed subjects						
Job-seeking motivation	T <sub>2</sub>	4.05	0.76	3.87	0.80	1.93(267)
	T <sub>3</sub>	4.02	0.77	3.96	0.84	0.42(149)
Job-seeking confidence	T <sub>2</sub>	5.04	1.02	4.58	1.21	3.36(265) <sup>c</sup>
	T <sub>3</sub>	4.55	1.32	4.61	1.29	-0.27(150)

<sup>a</sup>Nonparticipants include no-shows from the experimental group a sample of refusers (those who expressed no interest in participating in either the control or the experimental group).

<sup>b</sup>Unemployed and reemployed were coded respectively 0 and 1. Tests of significance are based on *t* test of these codes, which produce identical results to test of differences between proportions.

<sup>c</sup>*p* < .05.

<sup>d</sup>*p* < .01.

<sup>e</sup>*p* < .001.

program at the time of recruitment and are labeled "refusers." Prior analyses found the two subgroups to be very similar, and most importantly, to differ from the experimental participants group in the same way. Consequently, the data from all the nonparticipant subgroups were combined in the ensuing analyses.

In these analyses we sought to examine the following two questions: First, can nonparticipants be identified based on their psychosocial and demographic characteristics? Second, do nonparticipants show poorer outcomes than the participants who were shown to have benefited from the intervention with respect to achieving reemployment and mental health? A positive answer to these questions would suggest the need to intensify

recruiting and to identify the subgroups who should be targeted for the intensified efforts.

*T* tests were computed on the mean differences between the participant and nonparticipant groups. The results of the analyses that focused on reemployment outcomes are presented in Table III.

Compared to the participants, the nonparticipants were younger (34 vs. 40 years of age),  $t(449) = 5.72, p < .001$ , less educated,  $t(447) = 2.11, p < .05$ , had lower incomes in the past year,  $t(411) = 2.94, p < .01$ , reported greater economic hardship,  $t(443) = -1.95, p < .06$ , and greater confidence in their job seeking skills,  $t(445) = -5.06, p < .001$ .

In terms of reemployment outcomes, a higher proportion of nonparticipants than participants became reemployed at  $T_2$ ,  $t(379) = -4.04, p < .001$ . This difference continued as a trend at  $T_3$ ,  $t(386) = -1.62, p < .10$ . However, the nonparticipants who became reemployed at  $T_2$  had lower monthly earnings than the participants,  $t(90) = 2.15, p < .05$ . Of those who remained unemployed, nonparticipants had lower confidence in their job seeking skills at  $T_2$  than participants,  $t(265) = 3.36, p < .001$ . The nonparticipants also showed a trend of having lesser motivation for job seeking at  $T_2$ ,  $t(267) = 1.93, p < .10$ .

Contrary to expectations, these results suggest that compared to the participants, the nonparticipants were in fact better off at both posttest periods with respect to becoming reemployed. Additional comparisons between these groups on the mental health outcomes failed to produce significant differences. Since nonparticipants are not at greater risk of unemployment or poor mental health than participants, any additional efforts to achieve higher rates of participation by persuading nonparticipants to take part in the intervention may not be justified on the ground of achieving equality of outcomes. Achieving equality can best be served by providing the intervention in more locations to those who select themselves to take part in it.

Finally, there remains the question of whether the difference in the achieved reemployment outcomes between the participants and the nonparticipants can be accounted for by differences in their psychosocial and demographic characteristics that were identified at  $T_1$ . Although the answer to this question may not suggest new or different recruiting policies, it may highlight the dynamics that contribute to reemployment.

To examine this last question we repeated the above tests of differences between the participant and nonparticipant groups using analyses of covariance. In these analyses, the Time 1 demographic and psychosocial variables that are listed in Table III were entered as covariates. The results of these analyses failed to modify any of the significant differences found before or to produce other differences. Thus, the obtained differences in



reemployment outcomes are not the results of the initial differences identified at pretest.

## DISCUSSION

The design of field experiments to test the effectiveness of social interventions too often only focuses on the restricted goal of providing internally valid demonstrations of the benefits of the intervention. Even when successful demonstrations are achieved, the implementation of the intervention may be questioned on grounds of external validity, or generalizability. That is, the conditions of the field experiment may not represent those that will prevail for the actual implementation of the preventive program.

Two related problems often arise in field experiments that impede the direct estimation of the impact of the experimental intervention in its field implementation as an intervention program. Both problems derive from the reality of partial or selective participation in the experimental group by those who were randomly assigned to it. In a more fundamental form, partial participation appears when a certain proportion of experimentally assigned participants decline to participate, either at the outset when invited, or, later by not participating in the intervention itself.

At the earliest stage of the implementation of many social interventions, potential participants are offered the opportunity to take advantage of the program. The offer to participate usually includes information on the potential benefits of the program. In deciding whether or not to take part in a program, the candidates weigh the potential benefits of the intervention against the real and potential costs of participation. The real costs involve the time and effort of participation. Potential costs include foregone opportunities.

Depending on the circumstances, the offer to participate is often accompanied by the requirement to share some costs of participation, for example, through registration or other fees, or in contrast, an offer of reimbursement for various costs incurred such as transportation. Participation is then the end result of a complex decision process whereby the various potential and real costs and benefits are considered by the candidate.

Since participation is based on the program's candidates decisions, the wisdom or quality of these decisions determine, at least in part, the usefulness of the preventive program. The most valuable programs are those that attract participants who derive the maximum benefit from their participation while omitting those who are least likely to benefit. Traditional assessments of intervention programs focus on the extent to which

participants benefited from the program by comparing them to a control group of nonparticipants. Obviously, a valid comparison requires that participation and nonparticipation be determined by a random procedure, not by candidates' choice. A more comprehensive assessment of the value of a social program must include an answer to the question of whether the program attracts the participants that (a) have the greatest need for improvement, and/or (b) are most likely to benefit from it. If the answer is negative, the value of the program can be enhanced by additional efforts to improve the recruitment of the more appropriate candidates.

When the experimental condition of a tested intervention includes a significant proportion of nonparticipants, special analyses are needed to determine the intervention's effects on those who actually participated, not just the entire experimental group. These analyses attempt to preserve the internal validity of the experimental design and at the same time provide an externally valid estimate of the treatment effects on participants. It is possible to provide such an estimate when certain conditions prevail in the study design and its execution. The most critical condition is the attempt to treat experimental subjects who did not show up, or dropped out of the experiment early, in exactly the same fashion as the control subjects who did not receive the experimental treatment, including collection of follow-up data. If care is provided to insure this requirement, then Bloom's (1984) procedure for estimating the effects of participation should be the method of choice.

However, since in all cases, an offer to participate would have been made to the nonparticipants but not to the control subjects, there is a need to consider the possibility of an interaction effect between offering of the treatment and a personalogical variable. As described in the Methods section, our procedures insured that only respondents who expressed no preference for the experimental or the control condition were recruited for randomization into these conditions. Consequently, it seems implausible that an interaction between offering and personalogical variable influenced our results.

Once the effects on participants are determined, the question of who is in greater need for an intervention, and who can benefit more, participants or nonparticipants, should be addressed. The answers to these questions have implications to allocation of resources for recruitment policies when the intervention is implemented. As noted by Bloom (1984), the usefulness of his method of estimating the effects of participation extends to many circumstances where social programs, agencies, or services prohibit the provision of treatment strictly on the basis of randomization. These circumstances arise when the number of applicants for the program exceeds the number of placements available, but when the administrators of the program insist upon using certain criteria to select applicants for the program. The researcher may then randomly assign applicants to experimental

and control conditions. The agency or program can continue to use their preferred method of selecting applicants for treatment. If the agency or program agrees to restrict their selection to those that were assigned to the experimental condition then Bloom's estimation method can be used. Those who are assigned to the experimental condition but are not selected for treatment should be considered nonparticipants in the same manner as those who become nonparticipants by their own choice. These two types of subjects can be pooled in the analyses as the experimental nonparticipants subgroup.

Using data collected in a field experiment of preventive intervention for unemployed persons (Caplan et al., 1989), this paper applied a procedure suggested by Bloom (1984) to provide estimates for the effects of the intervention on actual participants. Effect size estimates were two to three times larger for the actual participant group than for the entire experimental group on employment outcomes (e.g., earnings) and mental health (anxiety and depression).

Additional analyses compared the achieved outcomes of the participants and the nonparticipants. The results demonstrate that the nonparticipants achieved significantly higher levels of reemployment at posttests and did not differ significantly on all the other outcomes including mental health. Eligible participants appear to engage in effective self-selection processes into and out of the intervention. These self-selection processes drew into the intervention the persons who needed it most. Thus, in implementing interventions with this pattern of results, resources could be better used in replicating the intervention in new sites rather than in making attempts to recruit a larger proportion of participants.

Nevertheless, the self-selection processes are not well understood insofar as they were not identified and measured in the course of our study but remain a subject for future research. The question of whether the self-selection processes *out* of the intervention were also effective in allowing the nonparticipants to maximize their own opportunities and benefits as well could not be answered using our data. An answer to this question requires a greater understanding of the role of psychosocial and demographic variables that contribute to reemployment and psychological well-being, and also requires unique research data on people who initially refuse and are then persuaded to participate.

## REFERENCES

- Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8, 225-246.

- Caplan, R. D., Vinokur, A. D., Price, R. H., & van Ryn, M. (1989). Job seeking, reemployment, and mental health: A randomized field experiment in coping with job loss. *Journal of Applied Psychology, 74*, 759-769.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Haney, C. (1991). Enhancing social support at the workplace: Assessing the effects of the caregiver support program. *Health Education Quarterly, 18*, 4.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 45*, 153-161.
- Kessler, R. C., Turner, J. B., & House, J. S. (1988). The effects of unemployment on health in a community survey: Main, modifying, and mediating effects. *Journal of Social Issues, 44*, 69-85.
- U.S. Bureau of Labor Statistics. (1986, July). Current labor statistics: Employment data. *Monthly Labor Review*. Washington, DC: U.S. Bureau of Labor Statistics.
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*, 156-157.